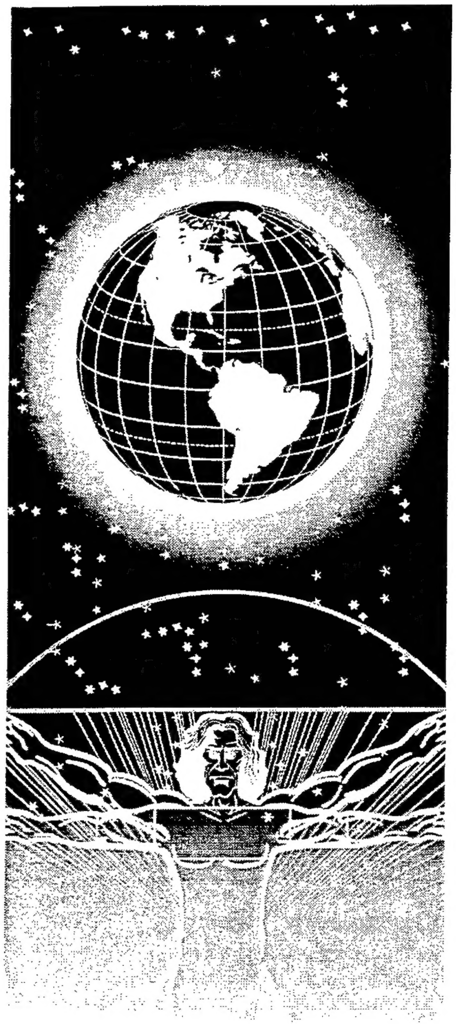
AFRL-HE-AZ-TP-2001- 0111



UNITED STATES AIR FORCE RESEARCH LABORATORY

DEVELOPMENT OF CRITERION MEASURES TO ASSESS RETENTION AND DECAY OF AEROSPACE PHYSIOLOGY KNOWLEDGE AND SKILLS

Cheryl Paullin

Amy F. McKee

Mary Ann Hanson

Jerry W. Hedge

Personnel Decisions Research Institutes, Inc.
43 Main Street SE, Suite 405
Minneapolis, MN 55414

October 2002

20031003 124

Approved for public release; distribution is unlimited.

AIR FORCE MATERIEL COMMAND
AIR FORCE RESEARCH LABORATORY
Human Effectiveness Directorate
Warfighter Training Research Division
6030 South Kent Street
Mesa AZ 85212-6061

NOTICES

WINSTON BENNETT JR
Project Scientist

HERBERT H. BELL
Technical Advisor

CURTIS J. PAPKE, Colonel, USAF
Chief, Warfighter Training Research Division

Copies of this report may be obtained from:

Defense Technical Information Center
8725 John J. Kingman Road, Suite 0944
Ft. Belvoir, VA 22060-6218
http://stinet.af.mil

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
October 2002

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*
June 1994 to June 1995

4. TITLE AND SUBTITLE
Development of Criterion Measures to Assess Retention and Decay of Aerospace Physiology Knowledge and Skills

5a. CONTRACT NUMBER
F41624-94-C-5014

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62202F

6. AUTHOR(S)
Cheryl Paullin, Amy F. Mckee, Mary Ann Hanson, Jerry W. Hedge

5d. PROJECT NUMBER
1123

5e. TASK NUMBER
A2

5f. WORK UNIT NUMBER
25

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Personnel Decisions Research Institutes, Inc
43 Main Street SE, Suite 405
Minneapolis, MN 55414

8. PERFORMING ORGANIZATION REPORT NUMBER
Institute Report No. 255

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Training Research Division
6030 South Kent Street
Mesa AZ 85212-6061

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL; AFRL/HEA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-HE-AZ-TR-2001-0111

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
Air Force Research Laboratory Technical Monitor: Dr Winston Bennett Jr., AFRL/HEA, (480) 988-6561 x-297, DSN 474-6297

14. ABSTRACT
This research examines the feasibility of developing criterion measures of Aerospace Physiology (AP) knowledge and skill retention and decay. We used two different criterion measures. One measure utilizes the situational judgment test methodology to create an inventory that measures the ability to apply AP knowledge and skills in situations actually encountered before, during, and after flight. We conducted several workshops with Air Force personnel to (a) collect descriptions of realistic situations in which physiological factors were affecting performance and (b) develop plausible response options for each situation. The second measure was a traditional job knowledge test consisting of multiple choice questions that measure recall of factual information related to AP. A small number of existing items were collected from quizzes administered during AP refresher training and new multiple choice items were written. Numerous obstacles were encountered when attempting to schedule workshops with Air Force personnel. This resulted in less comprehensive coverage of the AP content domain and fewer test items than were originally planned. Nevertheless, both measures show promise for assessing the retention and decay of AP knowledge and skills among Air Force aircrew members.

15. SUBJECT TERMS
Aerospace physiology; Criterion measures; FY94 SBIR; Knowledge retention; SBIR Phase I; Knowledge decay; Skill decay; Skill retention

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Ms Liz Casey
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED	45	19b. TELEPHONE NUMBER *(include area code)* 480.988.6561 x-188 DSN 474-6188

Table of Contents

List of Figures

Page

Preface

The work summarized in this report was performed by Personnel Decisions Research Institutes for the Air Force using funds provided through the DoD's Small Business Innovative Research (SBIR) program. This SBIR Phase I effort (Contract #F41624-94-C-5014) was completed in collaboration with the Skills Development Branch of the Armstrong Laboratory, Human Resources Directorate. The authors would like to thank Dr. Mark Teachout, Mr. Winston Bennett, and Dr. Richard Reinhart for their contributions to this research.

Summary

The goal of this Small Business Innovative Research (SBIR) Phase I effort was to examine the feasibility of developing criterion measures of Aerospace Physiology (AP) knowledge and skill retention and decay. Development of two different criterion measures was initiated. One measure utilizes the situational judgment test methodology to create an inventory that measures the ability to apply AP knowledge and skills in situations actually encountered before, during, and after flight. Several workshops were conducted with Air Force personnel to (1) collect descriptions of realistic situations in which physiological factors were affecting performance and (2) develop plausible response options for each situation. The second measure is a traditional job knowledge test consisting of multiple choice questions that measure recall of factual information related to AP. A small number of existing items were collected from quizzes administered during AP refresher training and new multiple choice items were written.

Numerous obstacles were encountered when attempting to schedule workshops with Air Force personnel. This resulted in less comprehensive coverage of the AP content domain and fewer test items than were originally planned. Nevertheless, both measures show promise for assessing the retention and decay of AP knowledge and skills among Air Force aircrew members.

Development of Criterion Measures to Assess Retention and Decay of Aerospace Physiology Knowledge and Skills

I. Introduction

One goal of this Small Business Innovative Research (SBIR) Phase I effort was to examine the feasibility of developing criterion measures to assess retention and decay of Aerospace Physiology (AP) knowledge and skills. These measures are one part of a comprehensive AP training needs analysis being conducted by the Skills Development Branch of Armstrong Laboratories/Human Resources Directorate (AL/HRTE) at the request of the USAF Surgeon General's Office. A second goal was to evaluate the commercial applicability of such measures.

Overview of Aerospace Physiology Training

In the early years of aviation, efforts to improve safety concentrated primarily on improving the reliability of aircraft mechanical systems. Over time, the reliability of aircraft mechanical systems has improved markedly but these systems have also become increasingly complex. As a consequence, the percentage of aircraft mishaps caused by human error has increased dramatically relative to the percentage of mishaps caused by equipment failure. According to Diehl (1989), human error has recently been estimated to play an important, contributing role in 50 to 90 percent of all aircraft mishaps.

The Air Force's (AF) AP training program is a critical component of efforts to improve aircrew reliability because skill decay in this area can lead to poor performance and increased likelihood of mishaps. AP is the study of how the human body and mind work in the flying environment (Reinhart, 1992). It is a term that encompasses a broad range of topics, all of which focus on how human performance can be affected by conditions encountered before, during, and after flight (e.g., diet and fitness, high altitude, G forces, spatial disorientation, and fatigue).

Given the recent reductions in manpower and budgetary resources in the military, it is extremely important for the Air Force to accurately determine training requirements so aircrew members receive training that is critical to mission performance without wasting scarce resources on unnecessary training. The AF Surgeon General's Office has identified a need to determine the most appropriate content and the necessary frequency of AP refresher training. Unfortunately, relatively little research has been conducted on the topic of AP knowledge and skill retention and decay. Teachout, Bennett, Barham, and Phalen (1993) reviewed three research projects that addressed the need for AP refresher training. They conclude that past research shows refresher

training to be beneficial, but there is considerable disagreement concerning the appropriate content and frequency of such training.

Currently, the AF medical community trains all aircrew members in AP. This training consists of an initial course that must be completed by all aircrew members prior to flying and a refresher training course that occurs at three year intervals throughout each aircrew member's military career (Teachout et al. 1993). The refresher training course includes academic (i.e., classroom) training and an altitude chamber ride. The academic portion of AP refresher training covers traditional physiological topics such as hypoxia, decompression sickness, visual illusions, diet and fitness, and may also cover less traditional topics such as situational awareness and/or crew resource management. The altitude chamber ride serves several purposes. First, it allows aircrew members to experience their personal symptoms of hypoxia (different persons experience somewhat different symptoms) and demonstrates dramatically how vision, thought processes, and motor coordination are affected by lack of oxygen. Second, it builds respect for the seriousness of the possible consequences when hypoxia occurs during flight. Third, it allows aircrew members to see how others are affected by hypoxia and thus makes it more likely that they will be able to detect when another aircrew member is experiencing hypoxia during flight.

A fairly standard AP refresher training curriculum exists but there appears to be a great deal of variability in the way the training is actually delivered. This variability is primarily a function of the amount of time spent on the various AP content areas, the extent to which the training content and materials are tailored to the needs of various weapons systems and/or crew positions, and the general quality of the instruction. It appears that most, if not all, AP instructors tailor their course(s) to particular types of aircraft and/or crew positions to some extent. For example, one refresher training course at the US Air Force School of Aerospace Medicine (USAFSAM) is tailored to the needs of aircrew members in Trainer, Attack, Reconnaissance, and Fighter (TARF) aircraft. Another refresher training course is tailored to the needs of aircrew members in Tanker, Transport, and Bomber (TTB) aircraft. In addition, during certain portions of the USAFSAM TTB training course, trainees are divided into two subgroups. During this time, aircrew members from the back of the aircraft (e.g., loadmasters) review egress procedures while aircrew members from the flight deck (e.g., pilots, flight engineers) review visual illusions and spatial disorientation. The topics covered in the TARF and TTB courses are fairly similar. However, the amount of time spent on various topics differs. For example, the topic of G-induced Loss of Consciousness (G-LOC) receives minimal coverage in the TTB course because this physiological effect rarely occurs in heavy aircraft. In contrast, G-LOC receives greater emphasis in the TARF course because this physiological effect is quite likely to occur in these types of aircraft.

In 1993, the AF Surgeon General's Office requested researchers at AL/HRTE to conduct a

comprehensive training needs analysis to determine what content should be covered in AP training and how frequently refresher training should be provided. AL/HRTE staff implemented a two-part approach to this needs analysis (see Teachout et al. 1993, for details). First, they surveyed a large number of aircrew members directly, asking them how frequently they use AP knowledge and skills and also asking for their recommendations about the frequency with which AP refresher training should be provided. Second, they hired Personnel Decisions Research Institutes, Inc. (PDRI) to explore the feasibility of developing measures of AP knowledge and skill retention and decay. Thus, the training needs analysis involves collecting aircrew members' informed opinions about their training needs and also involves an effort to develop criterion measures that can be used to directly evaluate their level of AP knowledge and skills.

As part of the SBIR Phase I contract described in this report, PDRI began to develop two different criterion measures: a traditional job knowledge test (which is referred to as a Job Knowledge Inventory, or JKI) and an innovative, written situational test (which is referred to as a Situational Judgment Inventory, or SJI). Before describing the development of these measures, we briefly review research findings relevant to these two types of criterion measures.

Relevant Research

One commonly-used and well-accepted method for assessing the effectiveness of, or need for, training is a job knowledge inventory (JKI). JKIs are straightforward to develop and administer. They require individuals to answer multiple-choice questions related to critical on-the-job knowledge, skills, and abilities. JKIs are particularly useful for measuring knowledge of technical information, such as that involved in the AP training course. Good JKIs representatively sample the content domain of interest. Thus, they provide an assessment of the degree to which individuals possess the factual knowledge covered in a training course and/or required to perform a job.

JKIs are clearly useful as predictors of job performance and training performance. Dye, Reck, and McDaniel (1993) conducted a meta-analysis of studies which used job knowledge tests to predict job performance or training success. They report a corrected mean validity of .45 for studies predicting job performance and .47 for studies of training success. They also report that validities were larger for high complexity jobs and also when job-test content similarity was high. Similarly, Hunter (1983) combined the results of fourteen validity studies and reports that job knowledge tests correlated .67 with work sample criterion measures and .40 with supervisory ratings of job performance.

JKIs have also been used as criterion measures. For example, JKIs were used as one of several

criterion measures for nine military occupational specialties in a large-scale selection and classification study conducted by the US Army (Project A; Campbell et al., 1990) and for four Air Force Specialties in the AF's Job Performance Measurement (JPM) project (Hedge & Teachout, 1986; Laue, Hedge, Wall, Pederson, & Bentley, 1992).

JKIs provide an excellent method for assessing fundamental knowledge about a content area. Individuals must possess fundamental knowledge before they can apply that knowledge effectively. One disadvantage of JKIs, however, is that they are not particularly useful for assessing skill in *applying* that fundamental knowledge in real-life situations. Other approaches have been used in past research to evaluate the degree to which persons can apply knowledge in real-life situations. These methods range from hands-on work sample tests to oral situational interviews to written situational judgment tests. Hands-on work sample tests are very costly to develop and administer. Situational interviews and written situational judgment tests are less costly to develop, yet still incorporate real situations which require application of knowledge.

The situational interview is one type of structured interview. Like other structured interviews, situational interviews consist of a standard set of questions with carefully developed procedures for evaluating answers. Situational interviews are composed of questions that pose hypothetical situations and ask interviewees what they would do in those situations. The situational interview is well-suited to the measurement of knowledge *application* in situations likely to be encountered on the job. Very realistic, hypothetical job situations that require the application of knowledges and skills can be developed and described by the interviewer.

The theoretical basis for the situational interview lies in Locke's (1968) model of goal setting which states that a person's intentions or goals are related to his/her subsequent behavior. Latham, Saari, Pursell, and Campion (1980) were the first to describe development and validation of a situational interview. These authors developed an interview scoring guide by asking experts to define, in behavioral terms, good, acceptable, and unacceptable answers to the interview questions. Interviewers were then trained to use these benchmarks as a guide in evaluating interviewee answers. Researchers at PDRI have developed a response checklist method for evaluating answers in a situational interview (Lammlein, Houston, & Paullin, 1993; Paullin, 1993; Paullin, Hough, & Dohm, 1991). In the response checklist method of evaluation, a comprehensive checklist containing possible answers is developed for each interview question. A scoring key is then derived by asking experts to judge how many points each possible answer should be worth. Interviewers simply match interviewee answers with the statements in the response checklists; they do not need to assign score values. Each answer receives the score value associated with the response checklist statement it most closely matches. Interviewers show good interrater reliability when either type of evaluation method is used. Bosshardt (1992) reviewed the available literature on situational interviews and reports a mean sample-size-

weighted interrater reliability of .79 across both types of evaluation procedures. Paullin (1993) reports interrater reliabilities which range from .75 to .97 across three studies that used the response checklist method of evaluation.

Structured interviews, including situational interviews, have often been used successfully as predictors of job performance. For example, Motowidlo et al. (1992) report an uncorrected validity coefficient of .22 as their best overall estimate of validity, across four independent studies. Bosshardt (1992) included only situational interviews in his review and reports a mean (uncorrected) sample-size weighted validity of .29. Structured interviews have received only limited use as criterion measures. However, Hedge and Teachout (1992) showed that an interview can be used successfully as a criterion measurement approach for AF jobs.

A structured, situational interview provides an innovative method for assessing retention and decay of AP knowledges and skills because it allows for the presentation of complex, true-to-life scenarios which simulate real-world AP-related situations. Responses to the interview questions can incorporate a series of behaviors that actually occur on the job. Most importantly, the interview questions can be used to assess the *application* of skills or knowledge. In addition, the response checklist scoring procedure allows for a greater degree of discrimination than the typical correct/incorrect scoring of a JKI. In other words, the response checklist can include answers that demonstrate various degrees of skill decay. Aircrew members whose skills have decayed to some extent, but not completely, are likely to provide answers in the moderate range of response checklist scores. The response checklists may also facilitate identification of which aspects of AP training decay more quickly (or more slowly) than other aspects.

One potential problem with the situational interview approach is that interviews are costly and time-consuming to administer. A trained interviewer must spend a substantial amount of time with each interviewee. Thus, situational interviews may not be practical for a large scale administration, such as that required to assess knowledge/skill decay for the number of aircrew members who attend AP refresher training.

A similar approach that is less costly and thus more practical is the situational judgment test approach. Situational judgment tests are in many ways similar to situational interviews, with the primary difference being that the situations are presented in a paper-and-pencil format. After reading the written description of a situation, respondents are asked to choose from a set of possible responses, which are also presented in a written format. While there is not a great deal of research on situational judgment tests available, researchers at PDRI and elsewhere have shown that situational judgment tests can be valid predictors of job performance (e.g., Forehand & Guetzkow, 1961; Motowidlo, Russell, Carter, & Dunnette, 1990; Phillips, 1992; Tenopyr, 1969). This type of test has most typically been used as a predictor of job performance, particu-

larly in managerial jobs. However, PDRI researchers have recently developed a situational judgment test which was used as a criterion measure for Project A. This test was designed to measure job knowledge in the supervisory aspects of soldiers' jobs, and research concerning the construct validity of this test provides support for its validity as a measure of supervisory job knowledge (Hanson & Borman, in press).

Additional information concerning the potential usefulness of situational judgment tests as criterion measures comes from research on written simulations, which are similar to situational judgment tests. Written simulations present respondents with written descriptions of realistic job situations. They differ from situational judgment tests in that they typically employ a branching format; each response leads to more information about the situation, and the respondent is again asked to choose between a new group of response alternatives. Written simulations have been used as measures of professional knowledge in several different fields, including law and medicine. Much of the available research on written simulations supports their construct validity as measures of professional knowledge and expertise. For example, comparisons of written simulation scores obtained by different groups of respondents have generally shown that when the groups are fairly distinct in terms of training and experience (e.g., students versus professionals), differences are significant and in the expected direction (e.g., Alderman, Evans, & Wilder, 1981; McGuire & Babbott, 1976).

PDRI researchers have developed an innovative method for scoring situational judgment tests that results in more reliable scores than simple "right/wrong" scoring. This approach is very similar to the response checklist approach, described earlier, that PDRI has used to score situational interviews. Briefly, subject matter experts are asked to rate the effectiveness of each response alternative for each item, and the mean effectiveness rating is computed for each alternative. Then, respondents are assigned the mean effectiveness rating associated with the alternative they identify as most effective as their item-level score. In order to obtain even more information from each test item, respondents can also be asked to indicate which response they believe is *least* effective in each situation. This could be seen as the ability to recognize and avoid very ineffective behaviors. By asking respondents to choose both the most and least effective behaviors, we may be able to examine the possibility that knowledge application deteriorates differentially depending on whether one must make a judgment about which action to take (i.e., which action is most effective) or which action to avoid (i.e., which action is least effective). A composite score can be computed for each respondent based on their choices for *both* the most and least effective responses. This composite has been found to be more reliable than either score alone (e.g., Hanson & Borman, 1990).

II. Approach

Formulate Plans

Our AP criterion development project began with a kick-off and planning meeting. The meeting was attended by representatives of PDRI and AL/HRTE. During this meeting we reviewed the project strategy as outlined in the research proposal. We also discussed how to obtain subject matter expert support from the Air Force for the workshops necessary to develop the two criterion measures. We agreed that AL/HRTE staff would schedule and coordinate the workshops with various active duty AF commands based on guidance from PDRI about the types and numbers of aircrew members required in each workshop.

We also collected Plans of Instruction (POIs) developed for the AP initial and refresher courses and refresher training materials available from USAFSAM and the AF Surgeon General's office. These materials provided a basic content outline that was used to guide development of the JKI and SJI.

Begin Development of a Situational Judgment Inventory (SJI)

Because the SJI was much more labor-intensive to develop than was the JKI, we began SJI development first.

Gather Situations. We gathered a number of situations directly from active duty and Air National Guard (ANG) aircrew members. In past research, we have found this to be the best method for generating realistic situations appropriate for development of an SJI. However, due to difficulties with the scheduling of AF personnel (a problem that plagued the entire research effort), we were unable to gather a sufficient number of situations directly from aircrew members. Therefore, we supplemented the pool of situations generated by AF personnel with situations adapted from other, existing sources (e.g., Class C physiological mishap reports and critical incidents available from other research).

A total of five situation generation workshop sessions were conducted at Brooks AFB, Randolph AFB, and the Minnesota ANG base. Twenty-four aircrew members participated in these workshops; twenty-one were active duty personnel and three were members of the Minnesota ANG. The participants included two females and three racial minority group members. We asked the AF to schedule aircrew members from a variety of weapon systems and a variety of crew positions to participate in the workshop sessions. We also made a special effort to include AP instruc-

tors and physiologists. Seven of the participants were physiologists or AP instructors (some of these persons also had flight experience). Of the remaining seventeen persons, thirteen had experience in heavy aircraft (i.e., TTB), one had experience in Fighter aircraft, and three had experience in rotary wing aircraft (i.e., helicopters). Seven different aircrew positions were represented, including pilot (8), navigator (3), flight engineer (2), loadmaster (1), boom operator (1), weapon systems officer (1), and flight nurse (1). All participants with flight experience had at least 500 hours of experience in their current weapon system and most had more than 3,000 hours.

At each workshop session, participants were asked to write brief descriptions of challenging and realistic situations that occurred before, during, or after flight, in which physiological factors affected performance. The situations were collected anonymously and participants were assured that the descriptions would not be used to identify persons who failed to file mishap or incident reports. Participants were then encouraged to describe events that had actually happened to them or to someone they knew and they showed no reluctance in describing such incidents.

Before writing these situations, participants were given detailed instructions concerning the characteristics of a "good" situation. For example, they were asked to focus on situations that require a thorough knowledge of AP and on situations in which an aircrew member's actions in response to a physiological problem could affect important outcomes (e.g., safety or performance). (See Appendix A for the instructions used in these workshop sessions.) We also provided participants with the basic outline of AP training content areas shown in Figure 1. Participants were instructed to write about situations that involved one or more of these AP content areas. There was some disagreement among workshop participants as to whether or not the content outline, and thus the situations being generated, should include the situational awareness content area. Some participants argued that situational awareness should not be a part of AP refresher training because it is covered in other training courses. Other participants pointed out that many AP refresher courses devote a significant amount of time to situational awareness. We compromised by instructing workshop participants to devote less effort to writing about situational awareness than about the other AP topics, but not to avoid situations that involved situational awareness (or the lack of it).

The 24 workshop participants wrote 124 usable situations. This is *substantially* fewer than we typically gather when developing an SJI and was not sufficient to proceed with development of an SJI. To increase the number of situations, PDRI staff used information from two additional sources to generate additional situations. First, we visited the Air Force Safety Agency (AFSA) to review mishap reports. We photocopied the narrative portion of approximately 100 Class C physiological mishap reports. It should be noted that no individuals or locations are identifiable in these narrative reports. In addition, AL/HRTE and PDRI staff agreed to keep all of the reports

TARGETED CONTENT AREAS

1. **Hypoxia/Hyperventilation**

2. **Trapped/Evolved Gases**

3. **Spatial Disorientation/Motion Sickness/Acceleration**

4. **Pressurization/Decompression**

5. **Vision/Night Vision/Visual Problems or Illusions**

6. **Hearing**

7. **Human Factors**
   - ♦ stress
   - ♦ self-medication
   - ♦ fitness/exercise
   - ♦ fatigue/circadian rhythms
   - ♦ diet/nutrition
   - ♦ alcohol, tobacco, and caffeine

8. **Environmental and Mission Stress**
   - ♦ smoke and fumes
   - ♦ noise and vibration
   - ♦ equipment
   - ♦ temperature extremes

9. **Situational Awareness**
   - ♦ channelized attention
   - ♦ task saturation
   - ♦ negative transfer
   - ♦ distraction
   - ♦ inattention
   - ♦ habituation

Figure 1. Aerospace Physiology Refresher Training Content Outline

in a secure location at all times. We were able to adapt 63 of these mishap reports into usable situations. Second, we reviewed critical incidents collected by other researchers at AL/HRTE. These critical incidents were not intended to target physiology and, in fact, most did not. However, PDRI staff were able to adapt 39 of the incidents into situations involving AP factors. Third, we reviewed situations that had been collected in another project PDRI is conducting for the Air Force (Hedge et al., 1993). In this project, an SJI is being developed to assess crew resource management (CRM) skills. Several situation generation workshops had already been conducted as part of this research effort. Although the focus of these workshops was CRM, aircrew members wrote some situations that were actually more relevant to AP than to CRM. Normally these situations would be discarded, but in this case PDRI staff were able to salvage 14 situations and use them to generate AP situations. Thus, we ended up with 240 situations. This number is somewhat misleading however, because we developed parallel versions of a number of the situations. The parallel versions involved the same basic scenario and physiological event, but each was written from the perspective of a different crew position or aircraft type. There were approximately 200 completely independent situations.

During the situation generation workshops we discovered that a number of physiologists, AP instructors, and active duty aircrew members believe that AP refresher training should be tailored to the needs of different types of aircraft and different crew positions. However, there are many different perspectives on exactly how much tailoring needs to be done. Everyone agreed that it seems impractical to develop a refresher course for each specific weapon system (e.g., F-16) and each crew position. The prevailing opinion seems to be that there should be at least three AP refresher courses, tailored to (1) multi-place, heavy aircraft (e.g., Tankers, Transports and Bombers; TTB), (2) single-seat and two-seat small, relatively fast aircraft (e.g., Fighters and Attack aircraft), and (3) rotary-wing aircraft. Even at this level of tailoring, some weapon systems do not fit neatly into any one category. For example, some Reconnaissance aircraft are multi-place, heavy aircraft while others are single-seat small, fast aircraft. Further, aircrew members in the back of multi-place aircraft have different needs than the flight deck crew in such aircraft. As described in the introduction of this report, the AP refresher courses offered at many AFBs have been tailored at least to the extent of distinguishing between heavy, multi-place aircraft (e.g., TTB) and smaller, faster aircraft (TARF). It is not clear which sessions aircrew members from rotary-wing aircraft attend.

In addition to these expert opinions, preliminary situation development revealed that a certain degree of tailoring automatically occurs, due to the nature of SJI items. SJI items must include enough information about the situation to allow respondents to choose among various responses. We learned that the following types of information are required, at a minimum, to be able to choose a response in most situations: (1) weather conditions, (2) mission type and duration, (3) crew position of the person who is the focus of the situation, and (4) presence or absence of

other aircrew members. Each of these pieces of information narrows the range of weapon systems and crew positions for which a situation can be relevant. For example, situations involving missions that last more than three hours are most relevant for heavy aircraft because other aircraft typically fly missions of much shorter duration.

In our original proposal, we planned to develop a single SJI to measure AP knowledge and skills. However, both expert opinion and initial SJI development suggested that more than one version of the SJI will be required.

At this point, we realized that we could devote a great deal of labor to gathering a comprehensive set of situations, a subset of which would be relevant to each weapon system and each crew position. Obviously, the increased effort would come at the expense of something else - namely, how far we could go in the SJI development process. After several discussions with AL/HRTE staff, we agreed to proceed with development of the SJI without gathering any additional situations. We felt this would allow us to achieve our Phase I goal of exploring the *feasibility* of an SJI for assessing retention and decay of AP knowledge and skills. At the same time, we recognized that we would not be able to develop comprehensive, tailored versions of the SJI for all weapon systems and crew positions. Nevertheless, we felt we could work toward a core set of items that would be relevant for most weapon systems and crew positions, as well as the beginnings of some tailored versions of the SJI.

Prior to the next set of workshops, PDRI staff reviewed and edited the situations to correct spelling and grammatical errors and to increase the clarity as much as possible. We also attempted to eliminate weapon-specific technical jargon while maintaining the structure and content of the situations. Twenty situations were eliminated because they were indecipherable and/or entirely redundant with another situation.

A few of the situations appeared relevant for all aircrew members (i.e., "common" situations) but many were only relevant for a subset of crew positions and/or weapon systems. We sorted the situations into overlapping subsets of situations we believed to be most relevant for various aircraft types and crew positions, creating four different subsets: (1) situations relevant to TARF aircraft, (2) situations relevant to TTB Flight Deck Crew positions (e.g., pilot, copilot, navigator, and flight engineer), (3) situations relevant to TTB Crew in the Back positions (e.g., loadmaster, boom operator, flight nurse), and (4) situations relevant to helicopters. Each subset included a common core of situations but each subset also contained unique situations.

Develop Response Options. Next, we conducted several workshop sessions to generate a set of viable response options for each situation. Thirty-seven active duty AF personnel participated in these response generation workshop sessions, which were conducted at Langley AFB.

Many of the participants were currently assigned to headquarters staff positions but all had a great deal of flying experience in at least one weapon system. All were male, three were racial minority group members, and two did not report their race. As with the situation generation workshops, we requested the AF to schedule personnel from a variety of weapon systems and crew positions. We also requested persons for whom it had been varying amounts of time since their last AP refresher training course. We did this to ensure that we would obtain some responses from persons likely to be at maximum levels of AP knowledge and skill decay. Workshop participants represented nine different aircrew positions: pilot (16), navigator (9), air surveillance technician (3), loadmaster (4), flight engineer (2), boom operator (1), communication system operator (1), and senior director - AWACS (1). Twenty-five participants had experience in multi-place, heavy aircraft (including large Reconnaissance aircraft such as the E-3 and E-4). Ten had experience in fighter aircraft or small, fast reconnaissance aircraft such as the F-4 and U-2. Finally, two had experience in helicopters. All but one participant had at least 500 hours of flight experience in their current or most recent weapon system and most had more than 1,000 hours of flight experience. About one-fourth of the participants had received AP refresher training within the past year, about one-fourth within the past two years, about one-fourth within the past three years, and about one-fourth had not received AP refresher training in more than three years (perhaps because they were no longer actively flying).

Participants were presented with a subset of those items we believed would be most relevant to them, based on their weapon system experience and crew position. They were asked to put themselves in the position of the person who was the focus of the situation and to write two or three sentences describing what they believed should be done to handle the situation. They were asked to provide sufficient detail to make it clear *why* their response would be effective. (See Appendix B for the instructions used in these workshops.) These workshop sessions resulted in between two and ten responses for each situation. In our past experience developing similar tests, we have found that at least 10 responses per situation are required to adequately cover the range of plausible responses. Thus, we once again were faced with a decision. We could schedule additional response generation workshops but this would prevent us from getting as far in the SJI development process as we had originally planned. The other option was to continue developing the subset of items for which we had the most responses. This would allow further exploration of the feasibility of the SJI as a criterion measure but would once again reduce the comprehensiveness of the resulting SJI. After several discussions with AL/HRTE staff, we decided to proceed with development of the SJI, focusing only on the subset of items for which we had the largest number of responses, even though we only had a marginally adequate number of responses for each of these situations. Because very few responses were available for situations relevant for helicopters or aircrew in the back positions, we decided not to conduct further development for these situations.

In preparation for the next set of workshop sessions, we edited the response options to clarify them, remove redundancies, correct grammatical and spelling errors, and remove weapon-specific technical jargon. Items were dropped from further consideration if (1) there was no variation in the responses, (2) the responses had nothing to do with knowledge of physiology, (3) the item was highly redundant with other items, or (4) the item was unrealistic. Almost half of the items were eliminated for one or more of the preceding reasons.

Review by Subject Matter Experts (SMEs). After experiencing significant scheduling problems and delays (due, in part, to units being placed on alert), we were able to conduct three SJI item review workshop sessions: one at Randolph AFB and two at Luke AFB. At this point, because we were approaching the end of the SBIR Phase I contract, AL/HRTE staff decided that we should focus only on development of a TARF version of the SJI. This version will contain situations relevant for all aircrew members (i.e., common core items) and situations relevant for small, fast aircraft (either single-seat or two-seat) that fly relatively short missions. We should note that this version of the SJI will not cover all situations relevant for TARF aircraft because we were unable to collect a comprehensive set of situations during the situation generation step of SJI development.

AL/HRTE chose to develop a TARF version because the TARF community appears to be more immediately faced with decisions about the content and frequency of AP refresher training than are the TTB or Rotary Wing communities. In addition, it appears that a single version of the SJI will be relevant for most, or all, TARF aircrew members. In contrast, it may be necessary to develop several versions of the SJI for TTB and/or rotary wing aircraft, each focusing on different crew positions (e.g., flight deck crew vs. crew in the back).

Twenty-three active duty AF personnel participated in the item review workshop sessions. Four were physiologists or AP instructors. The remaining 19 participants (all males, one racial minority group member) had flight experience in training aircraft (T-37, T-38), fighter aircraft (F-16, F-15), or both. Several were instructor pilots and several also had experience in heavy aircraft.

The SJI items relevant for TARF aircrew members were sorted into three subsets: (1) situations relevant *only* for single-seat or two-seat TARF aircraft, (2) situations relevant for all aircraft and all crew positions (i.e., common core items), and (3) situations relevant for flight deck crewmembers in either TARF or TTB aircraft. There were a total of 60 situations. Between three and six participants reviewed each subset. Workshop participants were asked to help clarify the wording of SJI items, remove redundant responses, add critical missing information, and add responses as needed. (Appendix C contains the instructions used in this workshop.) After the item review workshop, the SJI items were revised according to SME suggestions and an additional three items were eliminated. Figure 2 shows two example SJI items.

1. Exercise is one of your favorite activities and one of your personal goals is to run a complete marathon. However, you suspect that high intensity training for long distance running may be affecting your G-tolerance. What should you do?

   a. Balance your workout by including some weight training.

   b. Train for a triathlon instead of a marathon because the former requires more strength and muscle mass.

   c. Cut your running back to no more than five miles at a time.

   d. Stop training for a marathon.

   e. Long distance running should have little or no impact on your G-tolerance so if you are having problems with your G-tolerance, there must be some other cause.

   f. Train for a 10K run rather than a marathon.

2. You are a student pilot flying with an instructor pilot. The mission is progressing normally until you begin an accelerated stall. You experience severe disorientation so you transfer control to the instructor pilot. As he recovers to level flight, you start to sweat profusely. You feel very anxious. What should you do?

   a. Grab your airsickness bag in case you need to vomit.

   b. Look outside while the instructor pilot flies the aircraft.

   c. Gangload your regulator and try to control your breathing.

   d. Ask the instructor pilot to fly straight and level until you recover.

   e. Inform the instructor pilot of your problem.

   f. Evaluate your hypoxia symptoms.

Figure 2. Example Situational Judgment Inventory Items

Begin Development of a Job Knowledge Inventory (JKI)

In our proposal, we assumed that we could build a JKI based on items drawn from currently existing AP quizzes or exams. We expected to expend relatively little effort writing new JKI items. Therefore, we began the development of the JKI by searching for quizzes or exams developed by physiologists and/or AP instructors, assuming that such materials would be readily available. To our surprise, we found it difficult to locate such materials. Apparently, quizzes and exams are prepared at the discretion of individual physiologists and AP instructors. Because there is no requirement to test aircrew members' mastery of AP refresher course material, many physiologists and AP instructors do not develop or administer AP quizzes or exams at all.

We were able to obtain about 50 test questions from existing AP tests. A number of these items are "situational." In other words, the test question is immersed within a real-life AF situation. The difference between such "situational" JKI items and the SJI items is that for the JKI items there is one correct answer, out of four possible response options. For the SJI items, the response options (typically more than four) range in effectiveness from highly effective to ineffective, and no single response is scored as "correct." In addition, situational JKI items require examinees to recall a piece of information, whereas SJI items require examinees to apply their AP knowledge in response to a particular situation.

We supplemented the existing pool of JKI items by writing about 50 additional multiple choice test questions. These items are traditional multiple-choice items, similar to those found in most job knowledge tests. In other words, they are based directly on material included in relevant study materials and the item stems are written as a direct statement or question. Two example JKI items are shown in Figure 3.

If necessary, the JKI could be tailored to be more relevant for particular aircraft types and/or crew positions. Tailoring is easily done with a traditional JKI by developing a large pool of items, then selecting a subset of those items to achieve whatever type of tailored test is desired. For example, a JKI tailored for TARF aircraft may include a larger number of items testing knowledge of G-LOC than would be included in a JKI tailored for TTB aircraft.

The JKI can also easily be written to cover all AP content areas. This may not be the case for the SJI. First, some content areas are inherently difficult to measure using an SJI approach, because once the situation has been described, the "best" response is very obvious. Second, we have very few situations for some of the AP content areas due to the small number of participants in the situation generation workshops. The JKI, once fully developed, can be used to complement, or "fill in," any gaps in SJI content coverage.

1. Assume that following a rapid decompression to 35,000 ft you experienced NO physical or physiological symptoms. Your oxygen system worked correctly. What altitude do Air Force regulations dictate as the highest you may now fly while unpressurized?

    a. 10,000 feet

    b. 18,000 feet

    c. FL 250

    d. FL 350

2. How long do emergency and portable oxygen assemblies last?

    a. The emergency oxygen assembly lasts approximately ten minutes; the duration of the portable oxygen assembly varies.

    b. The portable oxygen assembly lasts approximately ten minutes; the duration of the emergency oxygen assembly varies.

    c. It depends on the user.

    d. It depends on the pressure setting and the altitude.

Figure 3. Example Job Knowledge Inventory Items

III. Conclusions and Applications of the Research

Conclusions

The SBIR Phase I effort to test the feasibility of developing criterion measures of decay and retention of AP knowledge and skills shows promise. In spite of difficulties in obtaining adequate numbers of AF SMEs, we were able to develop a small pool of plausible situations and responses options relevant for TARF aircrew members. We were also able to develop a pool of about 100 JKI items.

As noted above, we experienced extensive difficulty scheduling workshop sessions with AF personnel. In part, this difficulty was due to uncontrollable external factors (e.g., personnel being placed on alert). These difficulties were compounded when, in order to minimize travel, we requested to meet with relatively large numbers of AF personnel drawn from several AFBs near a single location (e.g., San Antonio). In several instances, it was quite difficult for units at nearby AFBs to provide the personnel resources requested. In the future, it appears likely that we will need to schedule smaller numbers of persons at a larger number of locations. This, of course, will require more extensive travel.

We also discovered that it is more difficult to develop an SJI for the AP content area than it is for many other content areas (e.g., supervisory knowledge). Some aspects of AP are of such paramount importance that emergency checklists have been developed specifically to help aircrew members deal with such situations. For example, if cabin pressure is lost (which involves the potential for hypoxia and decompression sickness), aircrew members simply follow the "rapid decompression" checklist. The only real knowledge required is realizing that cabin pressure has been lost and that the emergency checklist should be followed. For other aspects of AP, aircrew members may not need to understand the physiological causes of certain events (e.g., spatial disorientation) in order to effectively handle them when they occur. They simply have to deal with the symptoms. Even so, we found that, for at least a subset of the AP content areas, the SJI approach shows a great deal of potential.

Future Plans

Briefly, our short term plans include finishing the development of the TARF version of the SJI and creating a larger pool of JKI items. We then plan to conduct a large-scale field test of the TARF version of the SJI and of the JKI.

Finish the TARF Version of the SJI. We plan to conduct several additional workshop sessions with AF personnel from TARF aircraft, as well as physiologists and AP instructors. During these sessions, we will ask participants to judge the effectiveness of each response to each situation, using a 7-point rating scale. These effectiveness ratings will be used to develop a scoring key for the SJI. We will calculate the level of agreement among these SMEs for each response. Situations will be selected for the field test version of the SJI based on the following considerations:

- high agreement among SMEs concerning the effectiveness of the response options,
- the number of response options and the degree of variability in the effectiveness of the response options, and
- coverage of the AP content domain.

Create a Larger Pool of JKI Items. We recently learned that we may be able to contact a large number of physiologists and AP instructors to request any existing exam items. We hope that, through this process, we will gain a large number of potential JKI items. PDRI staff will review and edit the items for clarity, then conduct several workshop sessions with physiologists and AP instructors to review the items. During these same workshop sessions, we will ask physiologists and AP instructors to write new test questions, particularly for content areas that are lacking in coverage in our item pool at that time.

Field Test the SJI and JKI. Once we have fully developed a TARF version of the SJI and have generate a large pool of JKI items, we plan to conduct a large-scale field test of these inventories. We anticipate administering the SJI and JKI to a large sample of active duty, ANG, and/or Air Force Reserve personnel. The SJI will be administered only to persons who are currently, or have recently been, assigned to TARF aircraft. The JKI will be administered more broadly, because it can be constructed to contain items relevant for many types of aircraft and many crew positions.

Recommendations for Future Development

We recommend that the AF conduct additional development work on the SJI. This work would involve developing a comprehensive set of situations and response options for the TARF version, as well as situations and response options for several other versions (e.g., TTB Flight Deck Crew Positions, TTB Crew in the Back Positions, Helicopter Crew Positions). We also recom-

mend trying a more focused approach to SJI development, with smaller AF SME requirements for each workshop session. In this approach, SJI items would be "built" by small groups of AF personnel. During a single session, a small number of workshop participants would target only two or three AP content areas. The participants would be carefully selected to represent the appropriate aircraft type(s) and crew position(s). During the same session, plausible response options would also be generated. Thus, situation generation and response generation would be combined into a single meeting, rather than taking place during two independent sessions. Each session would produce only a small number of SJI items but those items would be more likely to tap the targeted knowledge and skills. We believe this approach would eliminate some of the obstacles that have hindered workshop scheduling during this SBIR Phase I effort because it requires smaller numbers of AF personnel at any one workshop session.

Commercial Applicability

We believe that these criterion measures may be of most interest to other branches of the military and to federal agencies involved in regulating the commercial aviation industry. To our knowledge, commercial airlines currently do not provide AP refresher training to flight crews, thus we believe they may be interested in this work as well.

During the Phase I SBIR research effort, we made initial contacts with the FAA and with several commercial airlines. The FAA seemed quite interested in this work. Persons from commercial airlines acknowledge the importance of AP knowledge among their flight crews, but also expressed concern that this work could lead to additional mandatory training requirements imposed by the FAA.

IV. Summary

We collected evidence regarding the feasibility of developing two different criterion measures to assess retention and decay of AP knowledge and skills. Our research suggests that it is feasible to develop an SJI and a JKI to measure AP knowledge and skills for aircrew members from a variety of weapon systems and crew positions. It also suggests that several versions of the SJI will be required to meet all the needs of the AP training community. We developed one such version for TARF aircrew members, but even this inventory would benefit greatly from further developmental efforts. Further, we learned that, in the future, it may be necessary to schedule only small numbers of AF personnel at any one location to provide input during SJI and JKI development workshops. This approach is certainly possible, although it will require more extensive travel.

References

Alderman, D.L., Evans, F.R., & Wilder, G. (1981). The validity of written simulation exercises for assessing clinical skills in legal education. *Educational and Psychological Measurement, 41*, 1115-11126.

Bosshardt, M.J. (1992). *A review of historical trends, variables and processes affecting interview outcomes, and new formats*. Minneapolis, MN: Personnel Decisions Research Institutes, Inc.

Campbell, C.H., Ford, P., Rumsey, M.G., Pulakos, E.D., Borman, W.C., Felker, D.B., de Vera, M.V., & Riegelhaupt, B.J. (1990). Development of multiple job performance measures in a representative sample of jobs. *Personnel Psychology, 43*, 277-300.

Diehl, A.E. (1989). Human performance aspects of aircraft accidents. In R.S. Jensen (Ed.), *Aviation psychology*. Brookfield, VT: Gower Publishing Company.

Dye, D.A., Reck, M., & McDaniel, M.A. (1993). The validity of job knowledge measures. *International Journal of Selection and Assessment, 1*, 153-157.

Forehand, G.A., & Guetzkow, H. (1961). The administrative judgment test as related to descriptions of executive judgment behaviors. *Journal of Applied Psychology, 45*, 257-261.

Hanson, M.A., & Borman, W.C. (1990, November). *A situational judgment test of supervisory knowledge in the U.S. Army*. Paper presented at the 32nd Annual Conference of the Military Testing Association, Orange Beach, AL.

Hanson, M.A., & Borman, W.C. (in press). *Development and contruct validation of the Situational Judgment Test (SJT)*. Minneapolis, MN: Personnel Decisions Research Institutes, Inc.

Hedge, J. W., Hanson, M. A., Borman, W. C., & Bruskiewicz, K. (1993). *Examining the feasibility of developing a situational judgment test for Air Force pilots* (Institute Report #245). Minneapolis: Personnel Decisions Research Institutes.

Hedge, J.W., & Teachout, M.S. (1986). *Job performance measurement: A systematic program of research and development* (AFHRL-TP-86-31). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.

Lammlein, S.E., Houston, J.S., & Paullin, C. (1993). *Job analysis and promotional procedure development for the Tennessee Department of Safety*. (Institute Report No. 221). Minneapolis, MN: Personnel Decisions Research Institutes, Inc.

Laue, F. J., Hedge, J. W., Wall, M., Pedersen, L., & Bentley, B. A. (1992, October). *Job performance measurement system development process* (AL-TR-1992-0120). Brooks AFB, TX: Technical Training Research Division, Human Resources Directorate.

Latham, G.P., Saari, L.M., Pursell, E.D., & Campion, M.A. (1980). The situational interview. *Journal of Applied Psychology*, *65*, 422-427.

Locke, E.A. (1968). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance, 3*, 157-189.

McGuire, C.H., & Babbott, D. (1976). Simulation techniques in the measurement of problem solving skills. *Journal of Educational Measurement, 4*, 1-10.

Motowidlo, S.J., Carter, G.W., Dunnette, M.D., Tippins, N., Werner, S., Burnett, J.R., & Vaughan, M.J. (1992). Studies of the structured behavioral interview. *Journal of Applied Psychology, 77*, 571-587.

Motowidlo, S.J., Russell, T.L., Carter, G.W., & Dunnette, M.D. (1988). *Revision of the Management Selection Interview: Final report*. (Institute Report No. 156). Minneapolis, MN: Personnel Decisions Research Institute.

Paullin, C., Hough, L.M., & Dohm, T. (1991, October). *Development of a content-valid structured interview for the entry-level firefighter position in the Minneapolis Fire Department: Final report* (Institute Report No. 212). Minneapolis, MN: Personnel Decisions Research Institutes, Inc.

Paullin, C. (1993). Features of structured interviews which enhance perceptions of fairness. In R.D. Arvey (Chair), *Perceptions, theories, and issues of fairness in the employment interview*. Symposium conducted at the 101st Annual Meeting of the American Psychological Association, Toronto, Ontario.

Phillips, J.F. (1992). Predicting sales skills. *Journal of Business and Psychology, 7*, 151-160.

Reinhart, R.O. (1992). *Basic flight physiology*. Blue Ridge Summit, PA: TAB Books.

Teachout, M.S., Bennett, W.R., Jr., Barham, B., & Phalen, W.J. (1993). *Determining intervals for Aerospace Physiology refresher training: An approach for research.* Paper presented at the Annual Conference of the Military Testing Association, Williamsburg, VA.

Tenopyr, M. (1969). The comparative validity of selected leadership scales relative to success in production management. *Personnel Psychology, 22*, 77-85.

Appendix A

AIR FORCE AEROSPACE PHYSIOLOGY

SITUATION GENERATION WORKSHOP INSTRUCTIONS

PURPOSE OF THIS WORKSHOP

In this workshop, you will be assisting in the development of a new test that may help determine the optimal frequency at which Aircrew Physiological Refresher Training should occur. When developed, this test will assess application of Aerospace Physiology (i.e., flight physiology) knowledge and skills in realistic situations. You have been asked to participate in these workshops either because you are a trainer/briefer for aircrew flight physiology or because you are an aircrew member and thus have likely experienced a variety of situations related to flight physiology.

The type of test we are developing is called a Situational Judgment Inventory. In this test, aircrew members will be presented with written descriptions of realistic and difficult scenarios in which flight physiology factors are affecting or could affect crewmember performance. The scenarios will be like those aircrew members actually experience during flight.

After reading each scenario, aircrew members will be asked what they would do in the scenario. Below are two example Situational Judgment Inventory items.

Situational Judgment Inventory
Example Item #1

You are about to embark on an overwater mission. You are at a small, foreign airfield without LOX servicing equipment available. You notice your LOX is low, but at the minimum required by regulations. The weather is VMC the entire route. You determine that if you had a depressurization in flight, you don't have enough LOX to last, but if you fly below 10,000 feet, you won't need it for a depressurization problem.

a. Check why your LOX is low and if you are still losing it.

b. Take off as scheduled, fly below 10,000 feet, and get your LOX serviced at the next stop.

c. Cancel the flight. Even though you may not need LOX for a depressurization problem, you may need it for smoke and fumes in the cockpit.

Situational Judgment Inventory
Example Item #2

You're at the end of a long trip. It's night and the crew is ready to set up for the ILS approach. Your cabin altitude has held at 5500 feet. As a 45 year old navigator, you know you aren't as tolerant to these long days as you used to be. Checking your approach plates, you find the numbers difficult to identify and the CRTs and instruments appear slightly blurred. You should:

a. Continue the approach, assuming that other crewmembers will monitor your performance.

b. Turn the instrument and cabin lights up and get on oxygen.

c. Put on your reading glasses.

d. Let someone else do the approach.

Today we are asking for your help as a subject matter expert to generate descriptions of challenging and realistic scenarios related to Flight Physiology. In future workshops we will obtain realistic responses to these scenarios, but today we would like to focus on defining the scenarios.

We are confident you have encountered many challenging situations that you can tell us about, but we first need to give you some important guidance for writing the type of scenarios that will be most appropriate for a Situational Judgment Inventory.

HOW TO WRITE SCENARIOS

To write scenarios, think about events in which factors related to flight physiology affected crewmember performance. The scenarios can be ones that you personally encountered in flight, situations that you have seen another aircrew member encounter, or situations you have heard about. The following is a list of the types of scenarios that would be appropriate for this test and some of the characteristics of these scenarios. Try to think of:

* difficult situations faced by aircrew members that involve physiological factors or the effect of physiological factors on the ability to perform safely and effectively.

* situations in which an understanding of the physiological effects of flight makes a difference in safety and performance.

* situations in which the aircrew member's actions in response to a potential physiological problem affect important outcomes (e.g., safety, performance).

* situations that require thorough knowledge of flight physiology.

A list of content areas is provided on the following page. Try to write scenarios that related to each of these areas or topics. However, don't be limited by this list. If you can think of situations related to flight physiology that do not fit into one of these areas, please describe these situations as well.

A good situation to be included in the Situational Judgment Inventory should have the following characteristics:

* It requires a response from the aircrew member.

* It is challenging. That is, the appropriate response is not obvious to everyone. For example, a challenging situation might be one which causes subtle impairment.

* It is realistic.

* The problem may not be readily apparent. The respondent may have to determine what the problem is in order to solve it, rather than just being handed a problem and asked to identify a solution.

* It requires knowledge of flight physiology and the ability to apply it, even when under stress.

* There is an appropriate or best way to respond to the situation, or at least some responses are better than others.

* It is important. An aircrew member's response in the situation will affect one or more outcomes that are important for personal and crew safety.

* It provides sufficient detail to help the respondent make a choice between possible actions.

* A response to the scenario can be communicated in just a few sentences.

TARGETED CONTENT AREAS

The following are general content areas from Aerospace Physiological Training course outlines. We would like you to focus on these as you write scenarios today. Try to think of one or more scenarios in each of the content areas. A scenario may relate to only one content areas or it may relate to several content areas. Please indicate the content area(s) to which you think each scenario relates. If a scenario you write is specific to a particular aircrew position or aircraft type, please indicate that as well.

1. **Hypoxia/Hyperventilation**

2. **Trapped/Evolved Gases**

3. **Spatial Disorientation/Motion Sickness/Acceleration**

4. **Pressurization/Decompression**

5. **Vision/Night Vision/Visual Problems or Illusions**

6. **Hearing**

7. **Human Factors**
   - ♦ stress
   - ♦ self-medication
   - ♦ fitness/exercise
   - ♦ fatigue/circadian rhythms
   - ♦ diet/nutrition
   - ♦ alcohol, tobacco, and caffeine

8. **Environmental and Mission Stress**
   - ♦ smoke and fumes
   - ♦ noise and vibration
   - ♦ equipment
   - ♦ temperature extremes

9. **Situational Awareness**
   - ♦ channelized attention
   - ♦ task saturation
   - ♦ negative transfer
   - ♦ distraction
   - ♦ inattention
   - ♦ habituation

Appendix B

Response Generation Workshop Instructions

AIR FORCE SITUATIONAL JUDGMENT INVENTORY RESPONSE GENERATION WORKSHOP: INSTRUCTIONS

Background

In this workshop, you will be assisting in the development of a new inventory that may help determine the optimal content and frequency of Aerospace Physiology Refresher Training. When developed, this inventory will assess aircrew members' ability to apply their knowledge of Aerospace Physiology in realistic situations. You have been asked to participate in these workshops because you are a highly experienced aircrew member.

The inventory we are developing is called a Situational Judgment Inventory. When aircrew members take this inventory, they will read written descriptions of realistic and difficult scenarios in which flight physiology factors are affecting or could affect their own or another crewmember's performance. The scenarios will be ones that aircrew members have actually experienced before, during, and after flight. After reading each scenario, aircrew members will be asked to indicate which of several responses they would choose to effectively handle the situation described. Below is an example of the type of item that will be included in the Situational Judgment Inventory. The full range of possible responses to this situation is not yet available.

Situational Judgment Inventory
Example Item

You are the aircraft commander of a heavy aircraft that is about to embark on an overwater mission. You are at a small, foreign airfield without LOX servicing equipment. You notice that your LOX is low but at the minimum required by regulations. However, you determine that if you have a depressurization problem in flight, you won't have enough LOX to last the entire trip. The weather is VMC the entire route. What should you do?

a. Try to find out why your LOX is low and whether or not the aircraft is still losing it.

b. Take off as scheduled, fly below 10,000 feet, and get your LOX serviced at the next stop.

c. Cancel the flight. Even though you may not need LOX for a depressurization problem, you may need it for smoke and fumes in the cockpit.

[other possible responses]

Today's Workshop

Today we are asking for your help as subject matter experts to generate response alternatives for this inventory. In previous workshops, experienced aircrew members wrote descriptions of challenging and realistic scenarios that involve at least one physiological factor and/or require situational awareness. The booklet you are about to receive contains a subset of these scenarios.

Read each scenario carefully. Then, put yourself in the place of the person in that situation. How should you respond in order to deal effectively with the problem that is described in the scenario? Your response should be realistic and should provide sufficient detail to make it clear why it is effective. To the extent possible, the response should reflect knowledge of flight physiology and/or situational awareness.

In general, responses to each situation can probably be described in two or three sentences, but feel free to write longer descriptions if more detail is necessary to adequately communicate what should be done. Do your best to write a realistic response for each situation, even if the crew position described is not your own.

Crew Position. If the crew position is specified in the scenario, respond as if you are the person in that crew position, even if it differs from your current crew position. For example, if you are a flight engineer and the scenario says "you are a pilot," write down how you think the pilot should respond in that situation, not how the flight engineer should respond. If you have absolutely no idea how a person working in a different crew position should respond in a scenario, then make a note to that effect under the scenario, and go on to the next one.

If the crew position is not specified in the scenario, assume that it is the same as your current crew position. For some scenarios the best response might be the same, regardless of crew position. For other scenarios the correct response may differ depending on the position of the person who is the focus of the scenario.

Multiple Booklets. Some of the scenarios we have collected are relevant for all crew positions and all weapon systems. Others are relevant only to particular crew positions and/or particular aircraft types. Thus, we have created several different booklets. Each booklet contains some scenarios relevant for all crew positions and aircraft types and some scenarios relevant only to particular crew positions and/or aircraft types. One set of booklets is targeted toward the crew on the flight deck of TTB-type aircraft, and another booklet is appropriate for people working in the back of these same aircraft. Another set of booklets is designed for TARF crewmembers, and a final set of booklets is most appropriate for helicopter crewmembers. You will be given a booklet containing scenarios that are most appropriate for your aircraft type and crew position.

When you write down how an aircrew member should respond to each scenario, do your best to figure out the best response based only on the information provided. It may be that the correct response for some scenarios is to seek out additional information. If a scenario does not provide

a critical piece of information that you would need to know what to do next, please make a note as to what information is missing. For example, if your response would be different depending on whether the mission is being flown during the day or at night and the time of day is not provided in the scenario, then write in what time of day it is, and then provide what you feel is the best response based on the information you added.

Much of the success of the flight physiology Situational Judgment Inventory depends on the quality of your responses today. Please read each item carefully, and try to come up with the best response you can.

Thank you very much for taking the time to help us today.

Appendix C

Item Review
Workshop Instructions

AIR FORCE SITUATIONAL JUDGMENT INVENTORY
ITEM REVIEW WORKSHOP:
INSTRUCTIONS

Background

In this workshop, you will be assisting in the development of a new inventory that may help the Air Force determine the optimal content and frequency of Aerospace Physiology Refresher Training. When developed, this inventory will assess aircrew members' ability to apply their knowledge of Aerospace Physiology in realistic situations. You have been asked to participate in these workshops because you are a highly experienced aircrew member, a flight physiologist, or an Aerospace Physiology instructor.

The inventory we are developing is called a Situational Judgment Inventory. When aircrew members fill out this inventory, they will read written descriptions of realistic and difficult scenarios in which physiological factors are affecting, or could affect, their own or another crewmember's performance. The scenarios will describe situations that aircrew members have actually experienced before, during, or after flight. After reading each scenario, aircrew members will be asked to indicate which of four or five responses they would choose to effectively handle the situation. An example of the type of item that will be included in the Situational Judgment Inventory is shown below.

Situational Judgment Inventory
Example Item

You have an early morning show. It's 1:00 a.m., you can't get to sleep, and you're sore from playing softball this afternoon. What should you do?

a. Apply heat to the sore muscles.
b. Take aspirin or Tylenol for the muscle pain.
c. Take a muscle relaxant.
d. Do something relaxing to help yourself fall asleep.
e. Take a non-prescription sleep aid (e.g., Ny-Tol)
f. Take Restoril (a prescribed sleep aid).
g. Report to the flight surgeon before your flight to discuss the problem.

Purpose of Today's Workshop

In previous workshops, experienced aircrew members wrote descriptions of challenging and realistic scenarios that involve at least one physiological factor and/or require situational awareness. Other aircrew members generated possible responses to these situations by writing down what they would do in each situation. In this workshop session, you will review a subset of these scenarios and responses.

We are asking for your help as subject matter experts to review and refine the scenarios and response options that have been collected previously. We will discuss the scenarios and response options and revise them to meet the following criteria:

- the scenario provides enough relevant information for respondents to decide between the alternative courses of action,
- the response options are as specific as possible. Each response should involve a specific behaviors rather than a vague, general thought or objective,
- the response options provide enough detail to clarify which would be more and less effective,
- the response options adequately cover the range of plausible responses,
- there is at least one highly effective or "correct" response alternative for each scenario.

In addition, we will discuss how, or if, responses can vary depending on the type of aircraft or a person's crew position. Based on all of this discussion, we will collapse redundant response options and drop those that are confusing or unnecessary. Finally, where appropriate and necessary, we will develop additional response options.